Task 3.6: Comparative verification of NWP and Nowcasting with spatial verification methods

AWARE Report

Contact:

Michael Hoff
Deutscher Wetterdienst
Department FE 15
Frankfurter Straße 135
63067 Offenbach am Main

E-Mail: michael.hoff@dwd.de
Tel.: +49 (0) 69 / 80 62 -2753

What has been done so far?

- ➔ R-package for common neighborhood verification methods (FSS, minimum coverage, pragmatic approach, contingency table after Stein & Stoop 2019, etc.)
  - ➔ Ensemble-FSS from Neighborhood-Ensemble-Probability (NEP) fields after Schwartz et al. (2010)
  - ➔ Spatial-temporal smoothing recently implemented (Le Duc et al., 2013)
  - ➔ Package still under development
- ➔ R-package for object-based verification after Davis et al. (2009)
  - ➔ „Total Interest“ and „Median of Maximum Interest“
  - ➔ DWD In-house object identification with KONRAD3D (will become operational)
  - ➔ Comparison of object attributes: centroid distance, minimum boundary distance, area ratio, intersection area ratio (more beeing planned)
  - ➔ Fuzzy logic parameter individually adjustable
  - ➔ Median of Maximum Interest (MMI) as summary score as well as all individual total interest values are available for further processing
- ➔ All scores can be visualised and processed via R-Shiny application

What will be done in the future?

- Neighborhood-verification:
  - Further developments on R-Package
  - Provide some interpretations of spatial-temporal ensemble FSS (and others) to extract key features and make it more user friendly
  - Implement Reliability diagrams and ROC curves, etc.
- Object-based verification:
  - Try to switch to R-package „sf" (simple features) since it provides more functionality in simple calculations and plotting
  - Provide some statistical output of matched objects (i.e. those objects with a total interest > x, where $0 \leq x \leq 1$
- At the moment, a real-time test system of Sinfony is running (NWP, nowcasting, combination products, 20+1 members each), so there is a lot of new data for further verification tests
- Note: Since June 2020, there is a new colleague focussing on object-based verification only. This leads to some delay in further results because of induction period

SINFONY reference data May/June 2016

Deutscher Wetterdienst
Wetter und Klima aus einer Hand

Model COSMO-DE (for the last time, afterwards switch to ICON-LAM):

➔ Most important updates:

ID	LHN	3D-Radar-Assimilation	Microphysic	Turbulence length	Hydrometeor-update LETKF
01	✓	✗	1-Mom.	150	✗
04	✗	✓	2-Mom.	500	✓

Area-based Nowcasting:

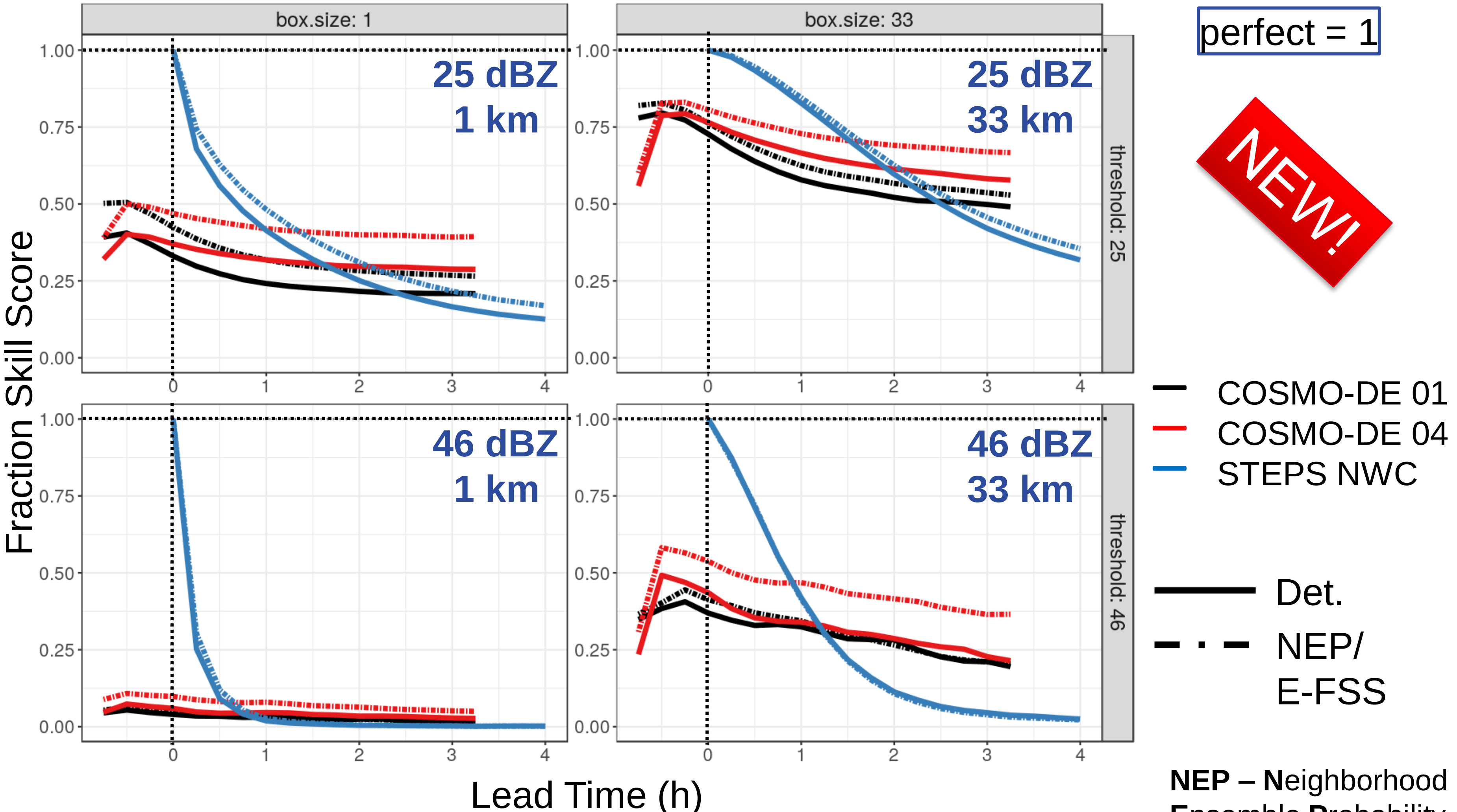
➔ Linear as well as **STEPS**-Approach for generating Nowcasting-Ensembles (Short-Term Ensemble Prediction System)

SINFONY

FSS (Deterministic + 20 mem. Ensemble)

Deutscher Wetterdienst
Wetter und Klima aus einer Hand

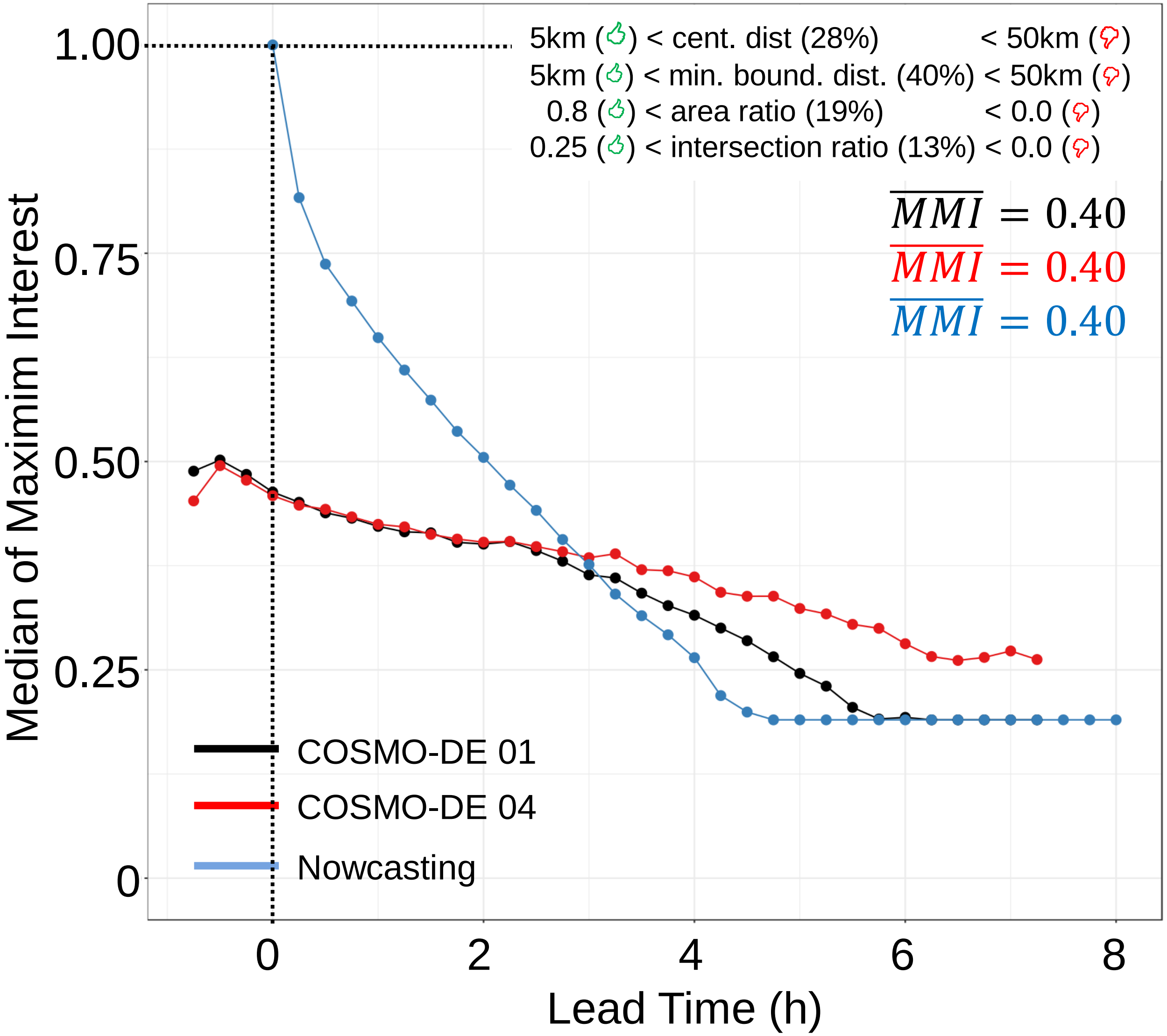
Obs.: RW, **Period:** 26.05.2016 – 25.06.2016, **Runs:** 10 – 15 UTC (hourly), **Lead time:** 0 – 4h



NEP – **N**eighborhood **E**nsemble **P**robability

SINFONY

MMI – Verification over entire Period



Interpretation of MMIF and MMIO (not shown):

COSMO-DE 01:
- less false alarms
- Too many misses

COSMO-DE 04
- massively too many false alarms during Spin-Up
- 0.25-4h: less false alarms, more misses
- 4-8h: less misses, more false alarms

Nowcasting
- Completely useless after 4h lead time
- less misses
- more false alarms

SINFONY

However…

…ist the comparison between observations and COSMO-DE really fair?

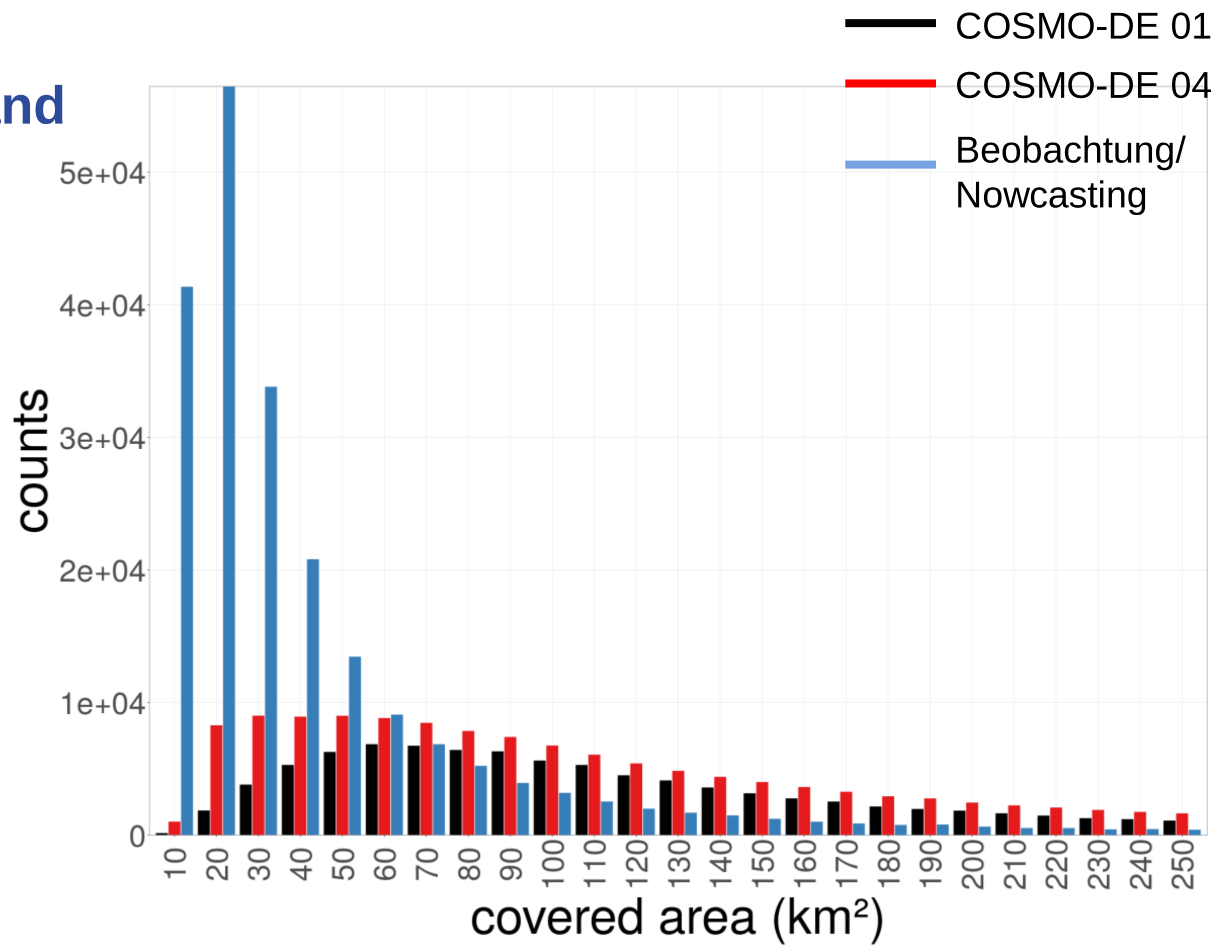
Too many small objects in the observation

Too many large objects in the forecast

→ Strong Bias in object attributes

→ What if we lower the threshold for the forecast only from 35dBZ to 30dBZ (slightly larger objects)

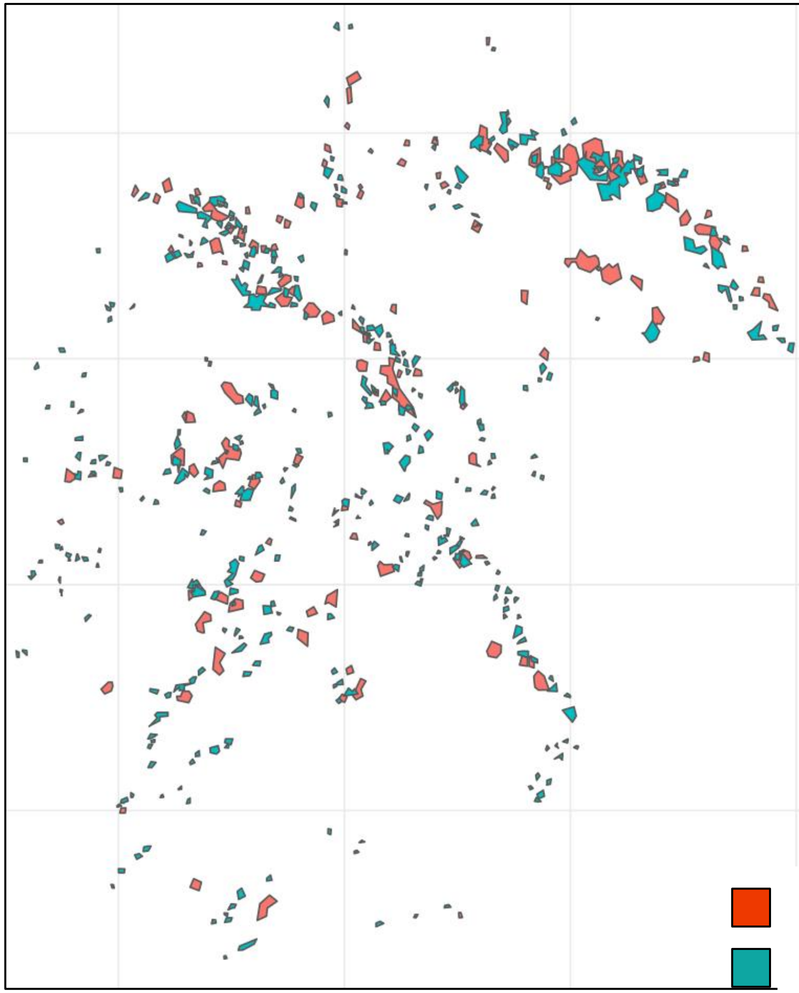
→ And exclude objects <50km² from the verification



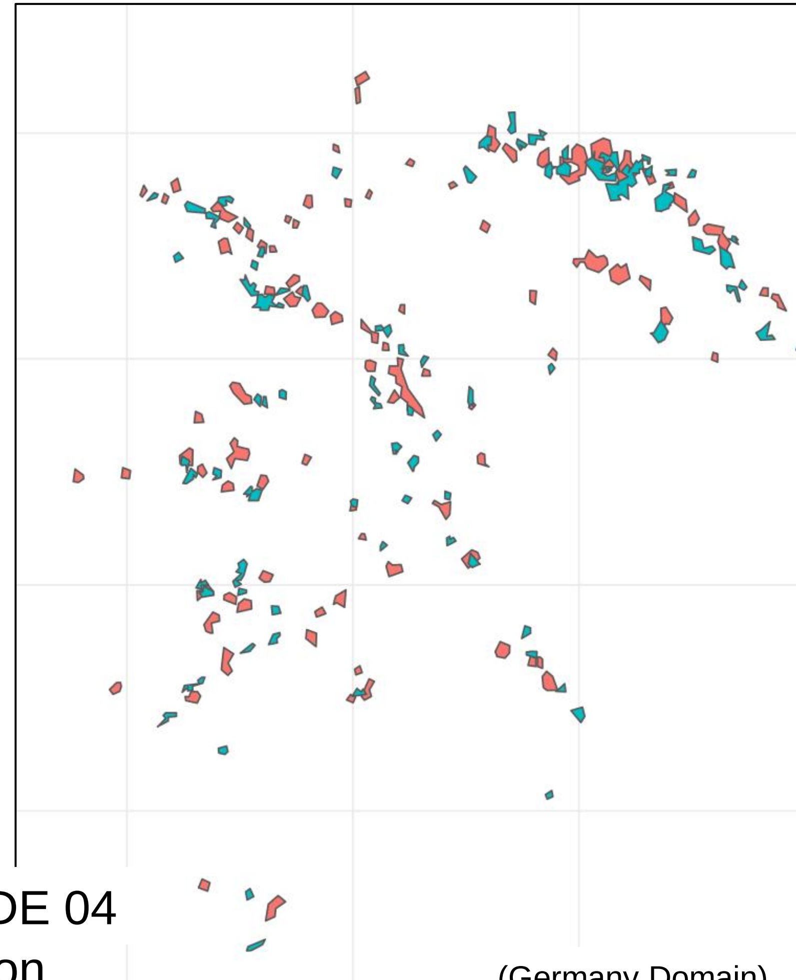
A qualitative Comparison

Deutscher Wetterdienst
Wetter und Klima aus einer Hand

All objects
30dBZ (Obs.) vs 35dBZ (COSMO-DE 04)



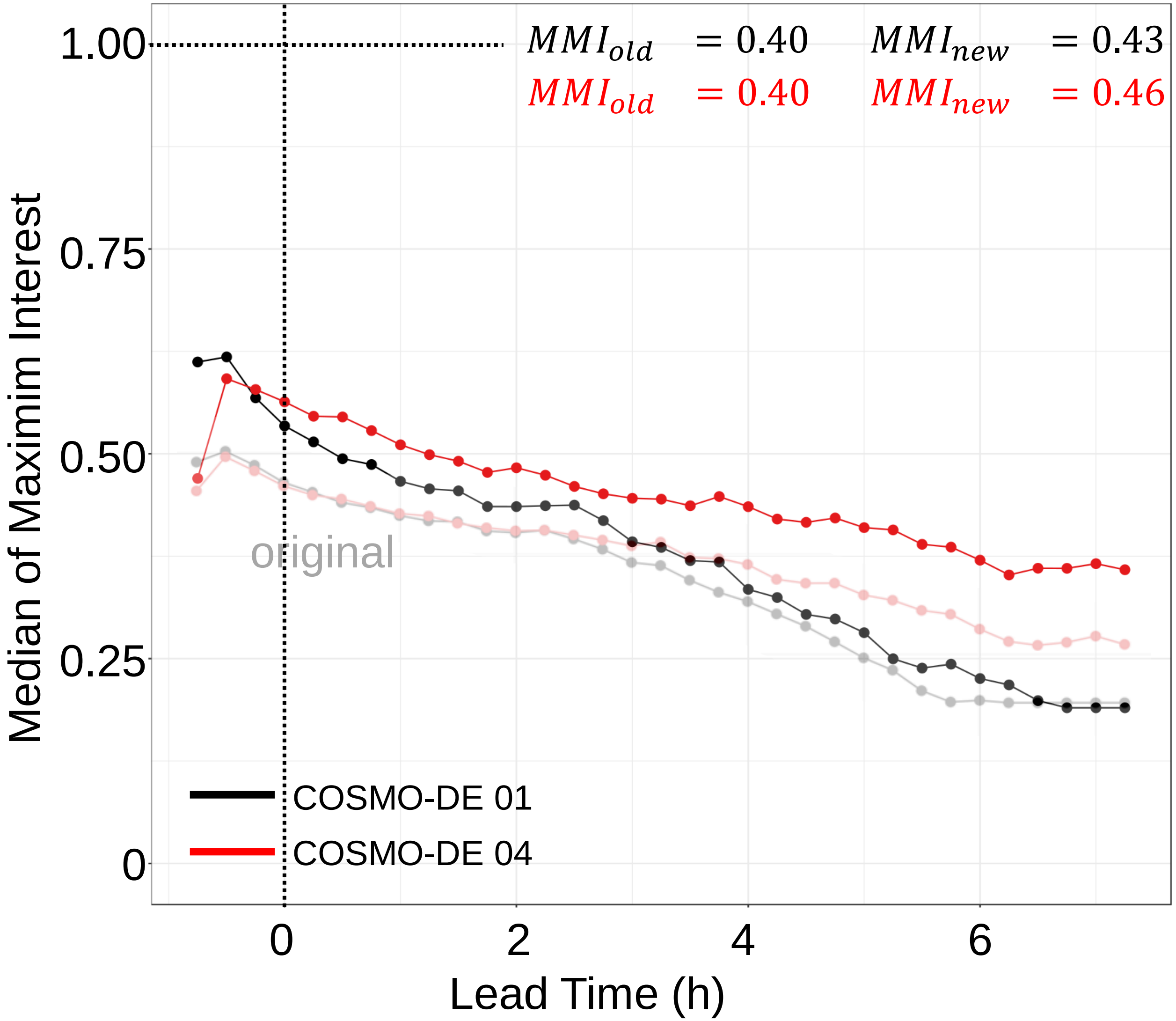
Objects > 50km²
30dBZ (Obs.) vs 35dBZ (COSMO-DE 04)



COSMO-DE 04
Observation

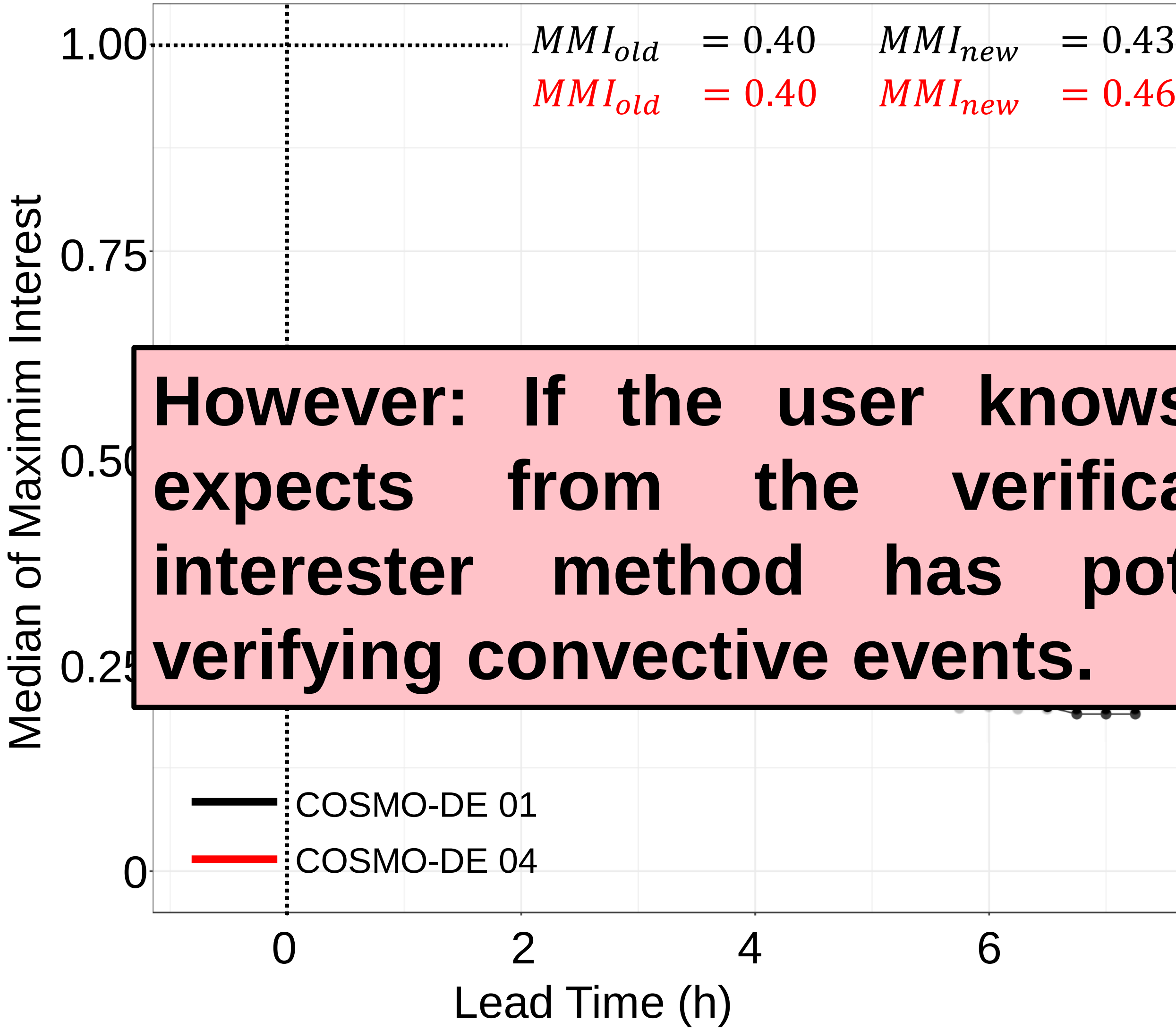
SINFONY

What says the MMI?



- The median affects the final score significantly
- Especially if there are too many small mismatched objects the score lowers drastically
- Bias correction might be an important aspect, especially when verifying convective events
- MMI quite sensitive in single cases, i.e. it should be interpreted over a longer time period

What says the MMI?

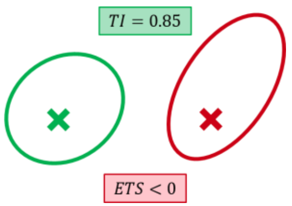


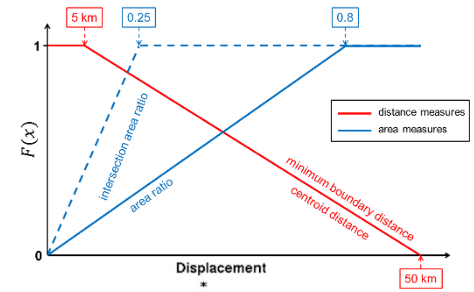
- The median affects the final score significantly
- Especially if there are too many small mismatched objects … rs
- … might … t … ally … convective events
- MMI quite sensitive in single cases, i.e. it should be interpreted over a longer time period

However: If the user knows what he expects from the verification, the interester method has potential for verifying convective events.

Conclusion object-based verification

- **Fuzzy logic** algorithm that compares different attributes of objects
- user-oriented verification: many parameter settings possible
- MMI as **„summary score“** but stratification on attributes possible
- **„Matching“** based on „total interest“ makes quantitative investigation of forecast errors of different attributes possible (e.g. displacement error)
- Object-based contingency tables (object-based skill scores)





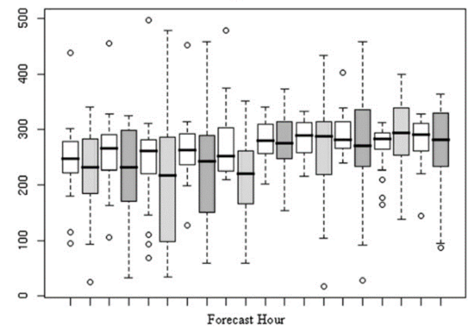


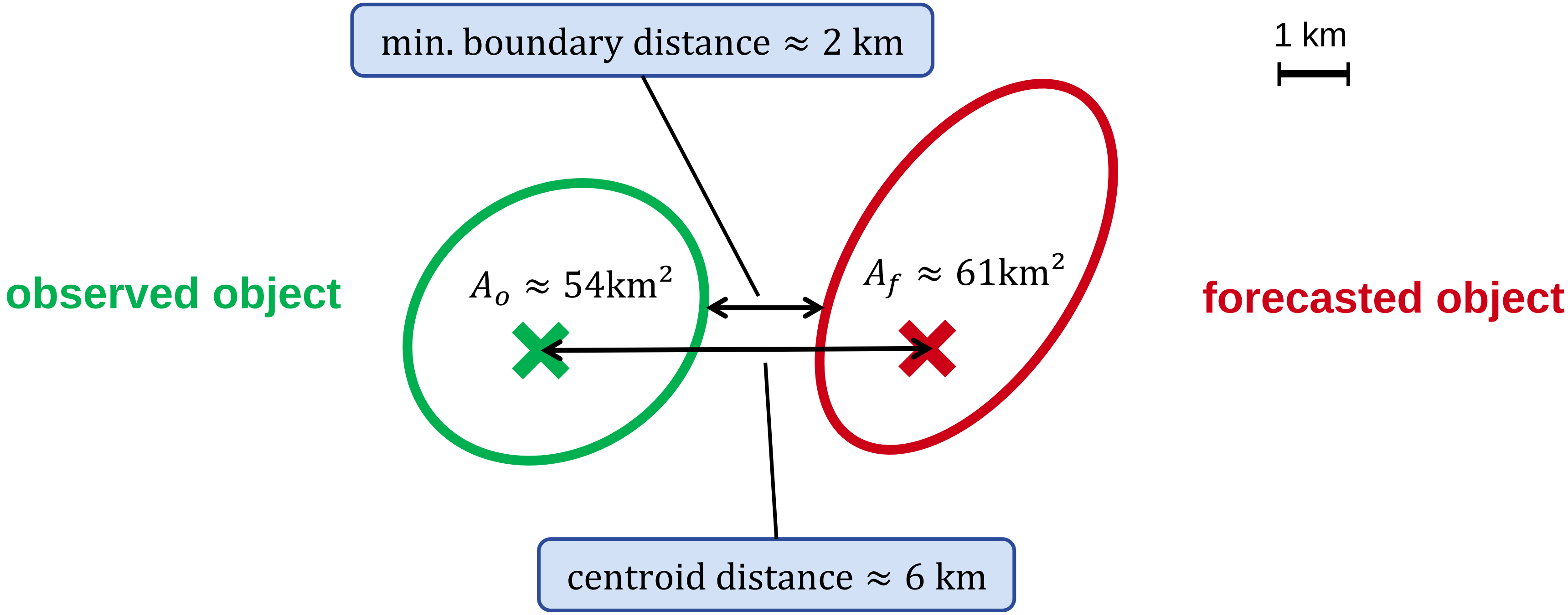
fig. from Gallus Jr. (2010)

<u>Unresolved issues:</u>

- Many parameter settings. Tuning necessary.
- MMI not defined, if no objects. False alarms and misses will not be punished (e.g. in case of linear nowcasting into calm night)
- Still unclear, how to **adapt** to **Ensemble forecasts**
- How to compare to other methods, e.g. neighborhood?

Appendix

Calculate the attributes of an object pair



$$\text{area ratio} = \frac{\min(A_o, A_f)}{\max(A_o, A_f)} \approx 0.88$$

$$\text{intersection area ratio} = \frac{A_{int}}{\text{mean}(A_o, A_f)} = 0$$

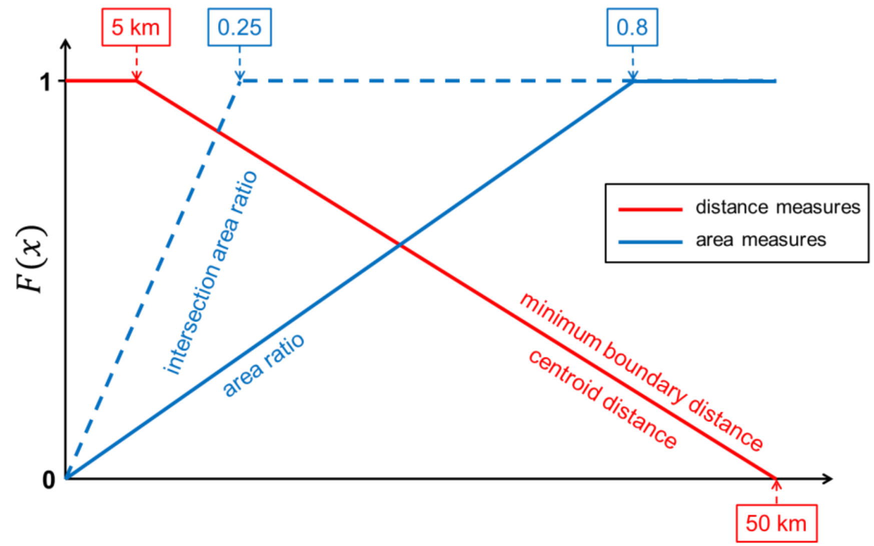
Build the interest for each object pair

Interester from the Method for Object-based Diagnostic Evaluation (MODE)
after Davis et al. (2009)

Weight *w* of attributes:

cent. dist (**28%**),
min. bound. dist. (**40%**),
area ratio (**19%**),
intersection ratio (**13%**)

Interest function *F*: includes limits for each attribute



Object pair *j*

$$I_{i,j} = c_i w_i F_{i,j} \quad [0,1]$$

Attributes *i*:

centroid distance,
minimum boundary distance,
area ratio,
intersection ratio

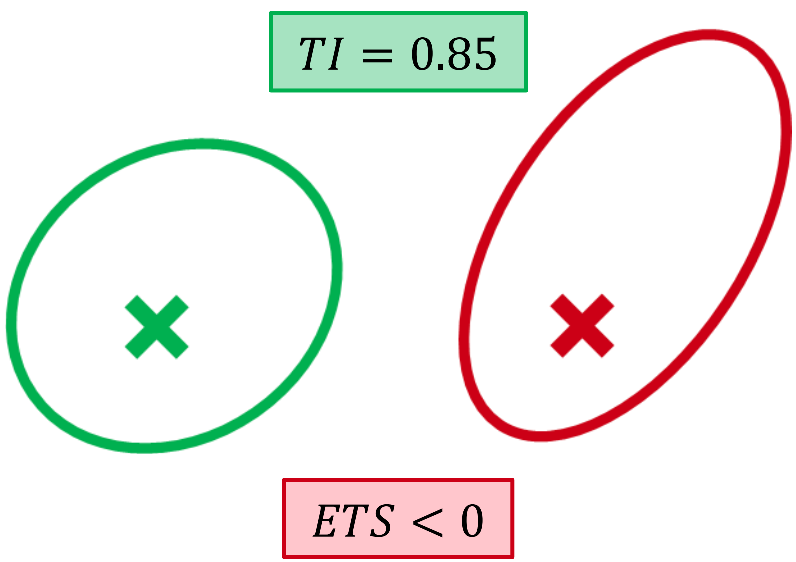
Confidence *c*: for each attribute

e.g. small centroid distance is less meaningful
if area ratio is greatly different → $c = \text{area ratio}$

Build total interest for each object pair

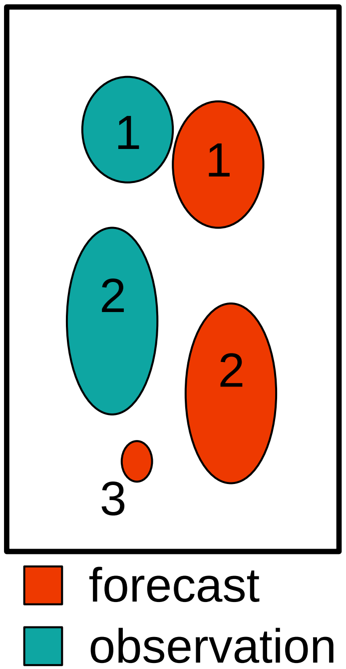
➔ sum up to „total interest“ for each object pair *j*

$$TI_j = \frac{\sum_{i=1}^{M} I_{i,j}}{\sum_{i=1}^{M} c_i w_i}, [0,1], M - \#\mathrm{attributes}$$



„Total Interest“ for all object pairs

$$TI_j = \frac{\sum_{i=1}^{M} I_{i,j}}{\sum_{i=1}^{M} c_i w_i}$$



“Total Interest” - matrix

observed objects

forecast objects	1	2	max:
1	0.90	0.75	0.90
2	0.50	0.80	0.80
3	0.40	0.55	0.55
max:	0.90	0.80	

0.80 (MMIF)

0.85 (MMIO)

False alarms lower the MMIF

missed objects lower the MMIO

$$\mathrm{median}(0.9, 0.8, 0.55, 0.9, 0.8) = \underline{0.8}$$

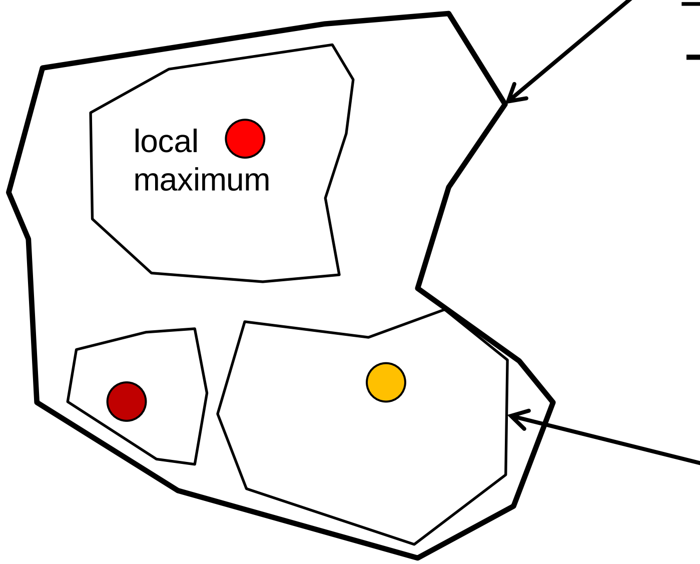
→ Median of Maximum Interest (MMI)

by Davis et al. (2009)

Identification of convective cells

In-House product: KONRAD3D

A **2D cell** for each radar & each elevation



Basic thresholding:

→ **here: 35 dBZ**

Adaptive thresholding:

A subcell has a min-max difference of at least **7 dBZ**

- **3D cells** by combination
- Applied to both, observations and simulations
- Note, the entire identification process is more complex